A K-sample Homogeneity Test based on the Quantification of the p-p Plot

Hinloopen, Jeroen; Wagenvoort, Rien; van Marrewijk, Charles

2008

TI 2008-100/1
Tinbergen Institute Discussion Paper

A *K*-sample Homogeneity Test based on the Quantification of the *p*–*p* Plot: The Harmonic Weighted Mass Index

Jeroen Hinloopen[1]
Rien Wagenvoort[2]
Charles van Marrewijk[3]

[1] *Amsterdam School of Economics, University of Amsterdam, and Tinbergen Institute.*
[2] *European Investment Bank, Economic and Financial Studies, Luxemburg;*
[3] *Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

A K-sample homogeneity test based on the quantification of the p-p plot: the Harmonic Weighted Mass index*

Jeroen Hinloopen†, Rien Wagenvoort‡,
and Charles van Marrewijk§

September 2008

Abstract

We propose a quantification of the p-p plot that assigns equal weight to all distances between the respective distributions: the surface between the p-p plot and the diagonal. This surface is labelled the Harmonic Weighted Mass (HWM) index. We introduce the diagonal-deviation (d-d) plot that allows the index to be computed exactly under all circumstances. For two balanced samples absent ties the finite sample distribution of the HWM index is derived. Simulations show that in most cases unbalanced samples and ties have little effect on this distribution. The d-d plot allows for a straightforward extension to the K-sample HWM index. As we have not been able to derive the distribution of the index for $K > 2$, we simulate significance tables for $K = 3, ..., 15$. An example involving economic growth rates of

*Thanks are due to Jan Kiviet, Robert Waldmann, Sanne Zwart, and seminar participants at the University of Adelaide, the University of Amsterdam, the University of Cyprus, and the Erasmus University Rotterdam for useful comments and constructive suggestions. The usual disclaimer applies.

†Corresponding author: University of Amsterdam, Amsterdam School of Economics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; J.Hinloopen@uva.nl; www.fee.uva.nl/io/jhinloopen.

‡Correspondence: European Investment Bank, Economic and Financial Studies, 100 Boulevard Konrad Adenauer, Luxemburg, Luxemburg; wagenvoo@eib.org.

§Correspondence: Erasmus University Rotterdam, Rotterdam School of Economics, Department of Economics, P.O. Box 1728, 3000 DR Rotterdam, The Netherlands; vanmarrewijk@few.eur.nl; www.charlesvanmarrewijk.nl.

the G7 countries illustrates that the HWM test can have better power than alternative Empirical Distribution Function tests.

Key words: EDF test, p-p plot, power, d-d plot.

JEL Classification: C12, C14

1 Introduction

To determine whether samples are drawn from the same distribution, Empirical Distribution Function (EDF) tests can be used if the underlying population distributions are not known. Examples include the Kolmogorov-Smirnov (KS) test (see Kolmogorov (1933) and Smirnov, 1939), the Fisz-Cramér-von Mises (FCvM) test (see Cramér (1928), Fisz (1960) and von Mises, 1931), the Kuiper (K) test (see Kuiper, 1960), and the Anderson-Darling (AD) test (see Anderson and Darling, 1952). Unfortunately the finite sample distribution is unknown for any of the concomitant statistics.

EDF tests quantify in one way or the other percentile-percentile (p-p) plots: the scatter plot of percentiles of two distributions for all entries of their joint support.[1] In this paper we introduce a new EDF statistic: the Harmonic Weighted Mass (HWM) index. It corresponds to the surface between the p-p plot and the diagonal, up to a scaling factor that depends on sample sizes. In case of two balanced samples absent ties we are able to derive the finite sample distribution of the test statistic. This is particularly useful for applications where samples are small, such as in experimental economics. Note that absent ties, in the limit the HWM index coincides with the L_1-version of the FCvM statistic. The HWM index differs from the FCvM statistic when there are ties in that it is invariant to the position of the tie in the sequence of order statistics. This makes it a more robust statistic.

We derive a formula with which the HWM index can be computed exactly under all circumstances. For that we introduce the diagonal-deviation (d-d)

[1] The obvious alternative is the scatter plot of order statistics of two distributions, the so-called quantile-quantile (q-q) plot. If distributions differ in scale and location only, q-q plots consist of straight lines, which is an attractive property when only differences in the shape of the respective distributions are of interest (Wilk and Gnanadesikan, 1968). But this is a drawback, of course, if differences in scale and location are to be revealed as well (see also Holmgren, 1995).

plot: the projection of the p-p plot onto the diagonal. Being able to compute the HWM index exactly allows us to simulate its distribution in all cases for which it is unknown. These reveal that in most cases the effect of ties on the distribution of the HWM index is negligible, and that the sample size correction factor deals adequately with unbalanced samples.

Extending the HWM index to the simultaneous comparison of $K > 2$ samples requires first the introduction of the K-dimensional p-p plot. The diagonal of this plot corresponds to the line that cuts all 2-dimensional spaces in equal halves. The K-sample HWM index then corresponds to the surface between this diagonal and the K-dimensional p-p plot. This surface is not uniquely defined however. An obvious norm is to consider for every point the shortest distance between the K-dimensional p-p plot and the diagonal. For this version of the K-sample HWM index we also derive a formula to compute it, relying again on the d-d plot. As we have not been able to identify analytically the distribution of the K-sample HWM index for $K > 2$, we provide significance tables up to $K = 15$.

None of the existing EDF tests dominates any of the other under all circumstances. Strictly speaking therefore, any sample comparison must involve the computation of all tests to rule out type-II errors as much as possible. The HWM test has more power than any of the other EDF tests when the p-p plot remains "close" to the diagonal over the entire probability space as it is the only EDF test that assigns equal weight to all distances between the respective distributions. It should therefore join the basket of EDF tests, also because in these cases a visual inspection of the p-p plot easily leads to incorrect conclusions. To illustrate this point the power of the HWM test is compared with that of the extended AD test (Scholz and Stephens, 1987) under circumstances where the latter is known to outperform existing alter-

natives. In particular, we compare the distribution of economic growth rates of all G7 countries where the distributions of two countries have significantly different tails. It appears that either test can have more power, depending on the subset of G7 countries that is considered.

2 The 2-sample HWM index

Consider the set of cumulative density functions $\Xi_1 = \{ F : \mathbb{R} \to [0,1] | \, \forall x, h \in \mathbb{R}$: $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$, $\lim_{h \downarrow 0} F(x+h) = F(x)$, and $a < b \implies F(a) \leq F(b) \}$. For F_1 and F_2 belonging to Ξ_1 the p-p plot depicts for every domain value z from the joint support of F_1 and F_2 the percentiles of one distribution relative to the other:

$$z \longmapsto \left[\begin{array}{c} F_1(z) \\ F_2(z) \end{array} \right]. \tag{1}$$

This is a plot in the 2-dimensional simplex. Written as a function rather than a plot it reads as:

$$p \longmapsto F_1(F_2^{-1}(p)), \; 0 \leq p \leq 1, \tag{2}$$

whereby $F_2^{-1}(p) = \inf \{ x : F_2(x) \geq p \}$. An example of a p-p plot for two continuous random variables is given in Panel a of Figure 1. Clearly p-p plots are mappings from $[0,1]$ onto $[0,1]$ and depict the correspondence between the two underlying distributions in probability space. They are key to the hypothesis underlying EDF tests:

$H_0 : \qquad F_1 = F_2.$

As the p-p plot coincides with the diagonal if, and only if, the two underlying distributions are identical, there are various well-known statistics to

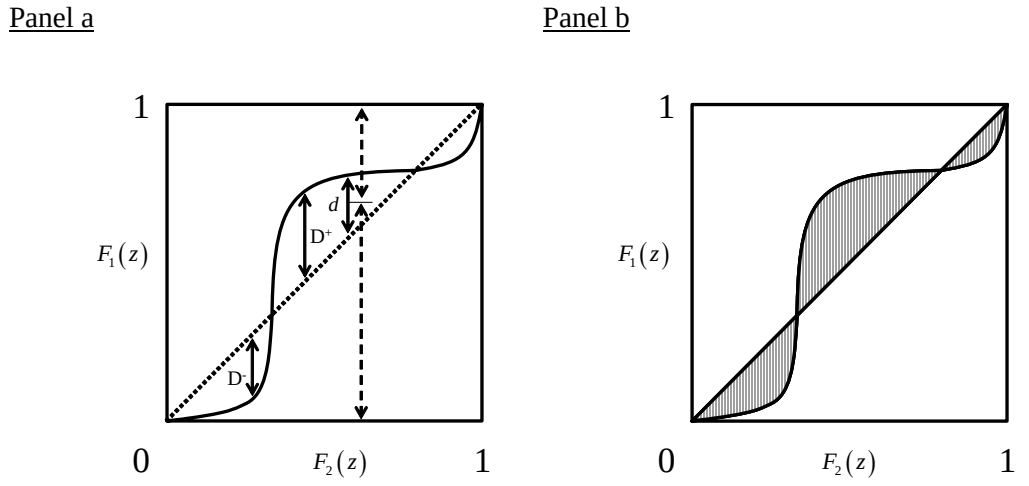

Figure 1: Theoretical p-p plot (panel a) and concomitant HWM index (panel b).

test H_0 that are based on the distance between the p-p plot and the diagonal. These include the Kolmogorov-Smirnov test (it considers the largest absolute value of the maximum positive distance, D^+, and the maximum negative distance, D^-; see Figure 1, Panel a), the Kuiper test (it considers the sum of D^+ and D^-), the Fisz-Cramér-von Mises test (it sums up over all squared distances d; see Figure 1, Panel a), and the Anderson-Darling test (it augments the FCvM-test by weighing every squared distance with the product of the distance between 0 and the centre of d, and the distance between 1 and the centre of d).[2] Although limiting distributions of these statistics are available (see Rosenblatt (1952), and Fisz, 1960), the finite sample distribution is not known for any of them.

[2]The area below the p-p plot corresponds to the Mann-Whitney-Wilcoxon U statistic (see Bamber, 1975). This statistic is used for testing whether one distribution first-order stochastically dominates another.

2.1 Definition

We propose the area between the diagonal and the p-p plot as the criterion for validating H_0 (see Panel b of Figure 1) as it is possible to derive under certain conditions the exact distribution of this area. Also, the related test can have better power than the existing alternatives. As the area reflects the extent to which the probability mass of the two underlying distributions is in harmony, we label it the Harmonic Weighted Mass (HWM) index.

The HWM index requires a continuous p-p plot. But for discrete random variables the p-p plot is also discrete. The continuous analogue of a discrete p-p plot is obtained by connecting the points of the discrete p-p plot through straight lines (Girling, 2000). Let X_1 and X_2 be two random variables with cumulative density functions $F_1(x)$ and $F_2(x)$ respectively. The coordinates of the resulting piece-wise linear continuous p-p plot can be defined as:

$$\begin{pmatrix} \widetilde{F}_1(x) \\ \widetilde{F}_2(x) \end{pmatrix} \equiv \left\{ \alpha \begin{pmatrix} F_1(z_{i-1}) \\ F_2(z_{i-1}) \end{pmatrix} + (1-\alpha) \begin{pmatrix} F_1(z_i) \\ F_2(z_i) \end{pmatrix} \right\},$$

where $\mathbf{z} \equiv \{z_1, ..., z_m\}$ are the ordered domain values of the joint support of X_1 and X_2, and where α is uniformly distributed over the unit interval. The 2-sample HWM index is then formally defined as:

Definition 1

$$HWM(F_1, F_2) \equiv S(n_1, n_2) \int_{-\infty}^{\infty} \left| \widetilde{F}_2(x) - \widetilde{F}_1(x) \right| d\widetilde{F}_1(x),$$

where the sample size correction factor $S(n_1, n_2)$ follows Rosenblatt (1952) and Fisz (1960) and reads as:

$$S(n_1, n_2) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}. \tag{3}$$

This sample-size correction factor speeds up convergence of the finite sample distribution towards the limiting distribution. Indeed, asymptotic critical values are adequate approximations of the exact critical values for samples containing as few as 8 observations. It also makes the analytical critical percentile values of the HWM index good approximations for unbalanced samples (see Section 2.3).

By virtue of the underlying p-p plot the HWM index is a non-parametric and distribution free mapping. It has two further properties (proofs of all properties are in Appendix A, Section 6.1):

Property P1 *(equality):*

$HWM(F_1, F_2) = 0 \iff \forall\ z \in [a, b] : F_1(z) = F_2(z).$

Property P2 *(order irrelevance):*

$HWM(F_1, F_2) = HWM(F_2, F_1).$

In practice the sample counterpart of a p-p plot needs to be considered. Let $X_1 = \{x_{1,1}, ..., x_{1,n_1}\}$ be n_1 realizations of the random variable X_1 with discrete sample CDF F_{1,n_1}, and let $X_2 = \{x_{2,1}, ..., x_{2,n_2}\}$ be n_2 realizations of random variable X_2 with discrete sample CDF F_{2,n_2}. The ordered set of values which either X_1 or X_2 assumes is denoted by $z_1, ...z_m$, whereby $m \leq n_1 + n_2$. In addition let z_0 and z_{m+1} denote $-\infty$ and ∞ respectively and define $\mathbf{z} \equiv \{z_0, ..., z_{m+1}\}$. The vertical coordinates of the discrete sample p-p plot are then defined by $P[X_1 \leq z_i]\ \forall\ z_i \in \mathbf{z}$ while the horizontal coordinates are given by $P[X_2 \leq z_i]\ \forall\ z_i \in \mathbf{z}$ (Bamber, 1975). Let $\widetilde{F}_{1,n_1}$ and $\widetilde{F}_{2,n_2}$ be the continuous counterparts of F_{1,n_1} and F_{2,n_2} respectively. The sample HWM index is then simply obtained by substituting in Definition 1 the population density functions by their sample counterparts. The resulting sample HWM index is consistent, which allows it to be used for testing H_0. That is:

Month	**1990**	**1993**
January	4070	4114
February	4204	3944
March	3885	3814
April	3866	3813
May	3808	3836
June	3854	3757
July	3762	3824
August	3786	3751
September	3764	3809
October	3872	3896
November	3926	3818
December	4292	4406

Table 1: Maximum water levels of the river Meuse at Borgharen Dorp, in millimeters, measured on every last day of every month. Source: Koninklijk Nederlands Meteorologisch Instituut, personal correspondence.

Property P3 *(consistency):*

$\lim_{n_1,n_2\to\infty} HWM\left(F_{1,n_1}, F_{2,n_2}\right) = HWM(F_1, F_2)$.

Accordingly, we also refer to the index value in Definition 1 as "the HWM index" if it is based on distributions of samples rather than population distributions.

As an illustration consider the water level of the river Meuse as it enters The Netherlands at Borgharen Dorp in 1990 and 1993. In 1993 the Southern part of Holland was plagued by severe floods and it is of interest to know whether the entire year 1993 was exceptional. Table 1 contains the maximum water levels in millimeters for both 1990 and 1993 recorded on each last day of every month. The resulting discrete sample p-p plot, its continuous counterpart, and the concomitant HWM index are all depicted in Figure 2.

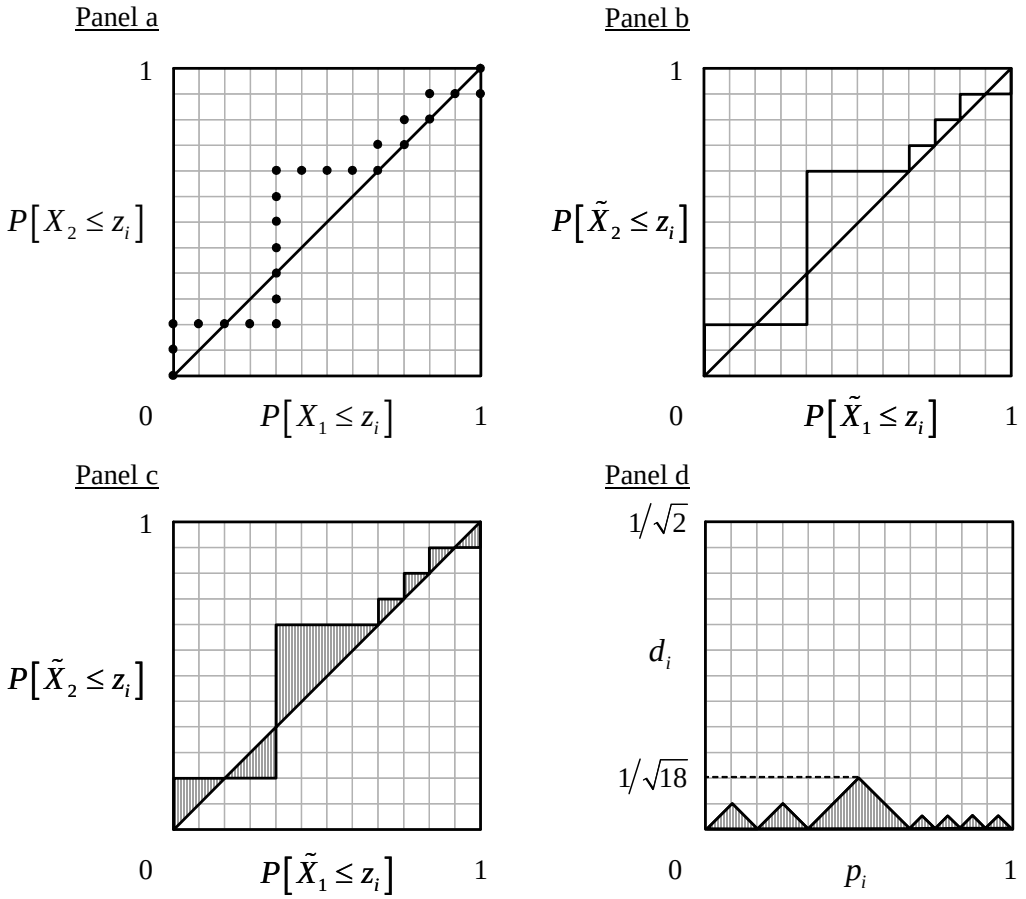

Figure 2: Discrete p-p plot for entering heights of the river Meuse into The Netherlands at Borgharen Dorp, 1990 versus 1993 (Panel a), the corresponding continuous p-p plot (Panel b), the concomitant HWM index (Panel c), and the resulting d-d plot (Panel d).

2.2 Computation

We have derived a method to compute the HWM index exactly for any $F_1, F_2 \in \Xi_1$. Note that the value of the shaded area in Panel c of Figure 2, which equals $\frac{7}{72}$, is straightforward to obtain. For samples with ties this is not necessarily the case. Within-sample ties impact on the number of coordinates that make up the p-p plot. As such they have no effect on the possible slope of the p-p plot. Between-sample ties may induce the p-p plot to have linear pieces with any positive slope. This complicates the exact computation of the HWM index.

For computing the HWM index we introduce the diagonal-deviation plot: the projection of the p-p plot onto the diagonal.[3] This means that each point $(F_1(z), F_2(z))$ on the p-p plot is projected on the average probability $p(z) \equiv (F_1(z) + F_2(z))/2$. Because the length of the projection vector equals $d(z) \equiv |F_1(z) - F_2(z)|/\sqrt{2}$, we arrive at the following definition:

Definition 2 *The diagonal-deviation plot is the projection of the p-p plot onto the diagonal:* $z \longmapsto \begin{bmatrix} p(z) \\ d(z) \end{bmatrix}$.

The d-d plot is obtained by first projecting the coordinates of the discrete p-p plot onto the diagonal. This gives the coordinates of the d-d plot: $D \equiv (p(z_i), d(z_i))$, $i = 1, 2, ..., m$, where $m \leq n_1 + n_2$, and $(p(0), d(0)) \equiv (0, 0)$. Next, linearly interpolate between these coordinates to obtain the d-d plot. The value of the HWM index then corresponds to the surface between the d-d plot line and the horizontal zero axis multiplied by the projection scaling factor $PS(2) = \sqrt{2}$ (see Panel d in Figure 2).

[3] Use of this projection is not necessary for computing the exact value of the 2-sample HWM index but it allows for a straightforward extension of the HWM index to the simultaneous comparison of $K > 2$ samples (see Section 3).

Computation of the HWM index requires the exact location of all intersections of the underlying p-p plot with the diagonal. Diagonal cutting points are all points $(F_1(z), F_2(z))$ for which $d(z) = 0$. Their location follows from the following lemma (the proofs of all lemmata are in Appendix A, Section 6.2):

Lemma 1 *When the p-p plot cuts the diagonal between points z_i and z_{i+1}, $z_i \in \mathbf{z}$, then*

$$p_h = F_{1,n_1}(z_i) + \frac{(F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i))(F_{1,n_1}(z_i) - F_{2,n_2}(z_i))}{(F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i)) - (F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i))}$$

if at probability p_h we have that $F_{1,n_1}(z_i) \geq F_{2,n_2}(z_i)$ and $F_{1,n_1}(z_{i+1}) \leq F_{2,n_2}(z_{i+1})$, i.e. the diagonal is "cut from below", and

$$p_h = F_{2,n_2}(z_i) + \frac{(F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i))(F_{2,n_2}(z_i) - F_{1,n_1}(z_i))}{(F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i)) - (F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i))}$$

if at probability p_h we have that $F_{1,n_1}(z_i) \leq F_{2,n_2}(z_i)$ and $F_{1,n_1}(z_{i+1}) \geq F_{2,n_2}(z_{i+1})$, i.e. the diagonal is "cut from above".

Obviously, the p-p plot can cut the diagonal at points that are not in D. Let L be the set with these diagonal cutting points consisting of l entries and let $I = D \cup L \equiv (p_i^*, d_i^*)$, $i = 1, 2, ..., m + l$ be ordered on p_i^*. The HWM index can then be calculated exactly (the proofs of all propositions are in Appendix A, Section 6.3):

Proposition 1

$$\begin{aligned} HWM(F_{1,n_1}, F_{2,n_2}) &= PS(2)S(n_1, n_2) \times \\ &\sum_{i=1}^{m+l} \left\{ \frac{(p_i^* - p_{i-1}^*)}{2} \left(2 \max\{d_i^*, d_{i-1}^*\} - \left|d_i^* - d_{i-1}^*\right|\right) \right\}. \end{aligned} \tag{4}$$

2.3 Hypothesis testing

Being able to compute the value of the HWM index exactly allows for the simulation of significance tables under varying circumstances. For a special case however the exact finite sample distribution of the HWM index is known.

2.3.1 Finite sample distribution

Consider the set of continuous distribution functions $\Xi_2 = \{F | \forall x,\ h \in \mathbb{R}$: $\lim_{x\to-\infty}F(x) = 0$, $\lim_{x\to\infty}F(x) = 1$, $\lim_{h\longrightarrow 0}F(x+h) = F(x)$, and $a < b \Longrightarrow F(a) < F(b)$, for $F(a), F(b) \in (0,1)\}$. Note that $\Xi_2 \subset \Xi_1$, that functions belonging to Ξ_2 are continuous and strictly increasing on their support, and that mass points are absent. In particular there are no ties. In what follows we derive the probability density function (pdf) of the HWM index under H_0 for distributions belonging to Ξ_2 and for two samples of equal size n. That is:

Assumption A1: $F_1,\ F_2 \in \Xi_2$.

Assumption A2: $n_1 = n_2 = n$.

Observe first some properties of the possible HWM index values:

Lemma 2 *Under A1 – A2, the number of possible distinct values of the HWM index is* $\Theta(n) \equiv 1 + n(n-1)/2$.

Lemma 3 *Under A1 – A2, the vector containing all possible distinct values of the HWM index is* $\mathbf{HWM}_n = \{HWM_n(1), ..., HWM_n(1 + n(n-1)/2)\}$, *where*

$$HWM_n(j) = \sqrt{\frac{n}{8}}\left(1 - \frac{2(j-1)}{n^2}\right),$$

$j = 1, ..., 1 + n(n-1)/2$.

The number of possible distinct outcomes of the HWM index varies quadratically with n. For example, $\Theta(20) = 191$ and $\Theta(100) = 4,951$. The number of possible distinct p-p plots increases much more rapidly, and goes from 1.38×10^{11} for $n = 20$ to 9.05×10^{58} for $n = 100$. Yet, the recursive pattern in which $\mathbf{HWM}_n$ varies with n allows us to identify its pdf for finite samples.

Let $BM(n) \equiv (\Delta_n | \, 2\Upsilon_n)$ for $n > 1$, where Δ_n is the upper triangular unit matrix of size n and Υ_n the unit vector of size n, and $BM(1) = [1\ 2]$, a 1×2 matrix (this matrix and all others introduced below are explained in Appendix A, Section 6.3). Further let $HM(1) \equiv 2$ and let $HM(n)$ be a $\Theta(n) \times n$ matrix with elements $hm_{jk}(n)$ defined as:

$$hm_{jk}(n) = \begin{cases} a_{(j-k+1)k}(n), & j = k, ..., \Theta(n-1) + k - 1, \ k = 1, ..., n, \\ 0 & otherwise, \end{cases}$$

where $A(n+1) \equiv HM(n)BM(n)$, whereby $A(1) \equiv 2$, and a_{jk} is an entry from the auxiliary matrix $A(n+1)$. Note that $A(n+1)$ is needed to construct $HM(n+1)$. Finally, let $\Omega(n) = \sum_i \sum_j a_{ij}(n)$ denote the number of all possible p-p plots and let $HM_j(n)$ denote row j of $HM(n)$. The following then holds:

Proposition 2 *Under* H_0 *and A1 – A2,* $HWM_n(j)$ *has probability* $p_n(j)$*, where*

$$p_n(j) = \frac{HM_j(n)\Upsilon_n}{\Omega(n)},$$

$j = 1, ..., \Theta(n)$.

The exact critical percentiles of the HWM index thus follow. They are given in Table 2 for sample sizes $n = 3, ..., 20$ at commonly applied significance levels.[4] High values of the HWM index imply a low probability that

[4] The entries do not display a monotonously declining pattern because the critical HWM index values are falling in sample size absent the sample size correction factor, whereas this factor itself is increasing in sample size.

percentile	90	95	97.5	99
n				
3	0.6124			
4	0.5303	0.6187		
5	0.5376	0.6008	0.6641	0.7273
6	0.5292	0.5774	0.6736	0.7217
7	0.5154	0.5918	0.6682	0.7445
8	0.5000	0.5938	0.6563	0.7500
9	0.5107	0.5893	0.6678	0.7464
10	0.5143	0.5814	0.6485	0.7379
11	0.5136	0.5911	0.6493	0.7462
12	0.5103	0.5784	0.6634	0.7485
13	0.5054	0.5808	0.6562	0.7467
14	0.4995	0.5804	0.6614	0.7424
15	0.5051	0.5903	0.6634	0.7486
16	0.5082	0.5856	0.6629	0.7513
17	0.4994	0.5801	0.6608	0.7516
18	0.5000	0.5833	0.6574	0.7500
18	0.4995	0.5849	0.6617	0.7471
20	0.5060	0.5850	0.6562	0.7510
∞	0.4993	0.5821	*	0.7518

Table 2: Exact critical values of the HWM index under A1 - A2 at percentile 90, 95, 97.5, and 99.

the underlying samples are drawn from the same distribution. For instance, for the data in Table 1 the HWM index value is 0.2381, which is substantially smaller than the critical percentiles at common significance levels. Although the southern part of The Netherlands was flooded in 1993, considering the entire year shows that the monthly maximum water levels of the river Meuse at Borgharen Dorp in 1993 were not significantly different from the levels in 1990.

2.3.2 Limiting distribution

In case of two continuous population distributions the HWM index converges asymptotically to the L_1-version of the FCvM statistic. In this case

no distinct points of the p-p plot exist which makes interpolation redundant. Schmid and Trede (1995) note that the limiting distribution of the L_1-version of the FCvM statistic corresponds to the limiting distribution of the L_1-norm of a Brownian bridge. Johnson and Killeen (1983) derive the analytical expression for the latter and tabulate its critical values. These values are in the last row of Table 2 (Johnson and Killeen (1983) do not include the limiting value for the 97.5^{th} percentile). Indeed, the critical values of the HWM index converge fairly rapidly to their limiting values.

2.3.3 Ties

Ties can be present in any sample, even for continuous population distributions due to rounding. Within-sample ties reduce the number of coordinates that constitute the p-p plot. Between-sample ties allow the p-p plot to remain closer to the diagonal. Ties therefore affect every EDF that quantifies a p-p plot. One way of dealing with ties is to use a randomized tie-breaking procedure (Dufour (1995), Dufour and Kiviet, 1998). Let U_i, $i = 1, ..., n_1+n_2$ be a random sample of $n_1 + n_2$ observations from a uniform continuous distribution. The observations $Z = X_1 \cup X_2$ can then be arranged following the order:

$$(Z_i, U_i) < (Z_j, U_j) \iff Z_i < Z_j \text{ or } (Z_i = Z_j \text{ and } U_i < U_j), \qquad (5)$$

which results in $n_1 + n_2$ different order statistics. The test statistic is then computed for these n_1+n_2 different order statistics rather then the $q < n_1+n_2$ order statistics from the original samples.

Alternatively the q order statistics are used and Proposition 2 is applied at the possible cost of a small size distortion in the critical area. Indeed, within-sample ties possibly increase the value of the HWM index whereas between-sample ties possibly reduce it. To assess the impact of ties on the

distribution of the HWM index we have to rely on numerical simulations because assumption A1 is violated.

As the influence of ties on the distribution of the HWM index turns out to be negligible in most cases, we report only on an extreme situation in that all observations constitute a tie.[5] Table 3 lists the size distortions of the critical percentiles implied by Proposition 2. This distortion is defined as $\Pr[HWM > cv] - cp$, where cv is the critical value of the HWM index under A1 – A2 and cp is the concomitant probability (i.e. 10%, 5%, 2.5% and 1%). Clearly what matters is the number of classes that underlie the sample CDFs. For instance, if there are only 2 classes, 0 and 1 say, the 90^{th} percentile is 0.400 while absent ties it equals 0.500, yielding a size distortion of -0.040. But size distortions fall rapidly when the number of classes increases. With as little as 7 classes the simulated 90^{th} percentile is already 0.485 and the size distortion is one percentage point only. It thus seems that for the HWM index the effect of within-sample ties and between-sample ties cancels out when the number of classes is sufficiently high. Because many applications will involve less than 100% ties (the only other study we know of that considers the effect of ties, Scholz and Stephens (1987), reports simulations up to situations where 60% of all observations constitute a tie), we conjecture that Proposition 2 is accurate under most circumstances where assumption A1 is violated.

Perhaps more importantly, the HWM index is less sensitive to ties than the L_1-version of the FCvM statistic. This is because the HWM index is invariant to the position of a tie whereas the FCvM statistic is not. For instance, let $X_1 = \{1, 2, 3\}$, and $X_2 = \{1.5, 1.5, 4\}$ with respective discrete CDFs $F_{1,3}$ and $F_{2,3}$. If the entries in X_2 are rounded upwards, yielding $\overline{X}_2 = \{2, 2, 4\}$ and $F_{2,\overline{3}}$, a between-sample tie arises at the second entry of both

[5] Simulations which show that the effect of ties on the distribution of the HWM index is negligible when the fraction of ties is smaller are available upon request.

percentile domain	90	95	97.5	99
{0, 1}	-0.047	-0.027	-0.018	-0.006
{0,.., 2}	-0.040	-0.022	-0.014	-0.006
{0,.., 3}	-0.023	-0.013	-0.007	-0.002
{0,.., 4}	-0.023	-0.012	-0.009	-0.003
{0,.., 5}	-0.018	-0.011	-0.005	-0.003
{0,.., 6}	-0.016	-0.009	-0.002	0.001
{0,.., 7}	-0.012	-0.007	-0.004	-0.002
{0,.., 8}	-0.011	-0.006	-0.003	-0.001
{0,.., 9}	-0.012	-0.007	-0.004	-0.001
{0,.., 10}	-0.008	-0.005	-0.003	-0.002
{0,.., 11}	-0.010	-0.008	-0.004	-0.001
{0,.., 12}	-0.006	-0.004	-0.004	-0.001
{0,.., 13}	-0.004	-0.006	-0.003	-0.002
{0,.., 14}	-0.005	-0.005	-0.003	-0.001
{0,.., 15}	-0.008	-0.006	-0.006	-0.002

Table 3: Size distortions of Proposition 2 due to ties in balanced samples. All samples consist of 50 entries which are drawn from the integer domains in the first column, where all entries have equal probability. The resulting simulated distribution for each row consists of 10,000 independent HWM index values.

samples. This gives $HWM(F_{1,3}, F_{2,\overline{3}}) = 0.2041$ and $FCvM(F_{1,3}, F_{2,\overline{3}}) = 0.1361$. Alternatively the entries are rounded downwards, giving us $\underline{X}_2 = \{1, 1, 4\}$ and $F_{2,\underline{3}}$. A between-sample tie now occurs at the first entry of both samples. This leaves the HWM index unaffected, $HWM(F_{1,3}, F_{2,\underline{3}}) = 0.2041$, while it does influence the FCvM statistic: $FCvM(F_{1,3}, F_{2,\underline{3}}) = 0.2722$.

2.3.4 Unbalanced samples

The second assumption underlying Proposition 2 is that samples are of equal size. To assess the influence of unbalanced samples we again have to revert to numerical simulations. Tables 6 through 9 in Appendix B (Section 7.1) list the simulated percentiles of the HWM index under H_0 for all possible unbalanced samples up to $n = 20$. Note the accuracy of these simulations in the extent to which the diagonal entries coincide with those in Table 2. It appears that the sample size correction factor (3) adequately deals with unbalanced samples. For larger samples the simulated percentiles suggest a simple rule of thumb: in case of unbalanced samples that are not in Appendix B take the analytical value of the HWM index for the largest sample size. Even in an extreme case where one sample consists of 3 entries while the other has 20 entries this yields a small approximation error.

3 The *K*-sample HWM index

For extending the HWM index to the simultaneous comparison of K samples we first have to introduce the K-sample p-p plot:

Definition 3 *The K-sample p-p plot depicts for every domain value z from the joint support of* F_1,..., F_K *the percentiles of one distribution relative to the others:* $z \longmapsto \begin{bmatrix} F_1(z) \\ ... \\ F_K(z) \end{bmatrix}$.

Obviously K-sample p-p plots are mappings from $[0,1]$ to $[0,1]^{K-1}$. Hypothesis H_0 then extends to:

$$H_0^* : \qquad F_1 = ... = F_K.$$

And the generalized assumptions A1 and A2 respectively read as:

Assumption A1*: $\quad F_1$,..., $F_K \in \Xi_2$.

Assumption A2*: $\quad n_1 = ... = n_K = n$.

3.1 Definition

The K-dimensional HWM index is defined as the surface between the K-dimensional p-p plot and the diagonal that cuts all 2-dimensional spaces in equal halves. For $K > 2$ this surface is not uniquely defined however. It depends on the point of the diagonal on which probabilities are projected. An obvious candidate is to take the shortest distance between the p-p plot and the diagonal, which implies that each point of the p-p plot is projected on the concomitant average probability.[6] For characterizing this distance we use the Mahalanobis distance:

Definition 4 $HWM(F_1, ..., F_K) \equiv PS(K)S(n_1, ..., n_K) \times$

$$\int_{-\infty}^{\infty} \sqrt{\sum_{i=1}^{K} \left(\widetilde{F}_i(x) - \widetilde{p}(x) \right)^2} d\widetilde{p}(x),$$

where $\widetilde{F}_i(x)$ is the continuous analogue of the possibly discrete CDF of X_i, $\widetilde{p}(x) = \frac{1}{K} \sum_{i=1}^{K} \widetilde{F}_i(x)$ is the average probability at x, where $S(n_1, ..., n_K)$ is a multi-sample scaling factor, and where $PS(K)$ is a factor that scales the

[6] We also considered alternatives, such as the minimum probability. This led to much more variability in the HWM index which, we think, makes it less suitable for hypotheses testing.

projection. In particular we generalize (3) to:

$$S(n_1...,n_K) = \frac{\left(\prod_{j=1}^{K} n_j\right)^{\frac{1}{K}}}{\left(\sum_{j=1}^{K} n_j\right)^{\frac{1}{2}}}. \tag{6}$$

And the projection scaling factor $PS(K)$ maintains the correspondence between the surface below the d-d plot and the surface between the K-dimensional p-p plot and the diagonal. It equals the length of the diagonal of the K-dimensional p-p plot:

$$PS(K) = \sqrt{K}. \tag{7}$$

Note that the 2-sample version of the HWM index in Definition 1 corresponds to Definition 4 with $K = 2$.

The properties of the 2-sample HWM index carry over to the K-sample version:

Property P1 *(equality):*

$HWM(F_1, ..., F_K) = 0 \Longleftrightarrow \forall\ z \in [a, b] : F_1(z) = ... = F_K(z)$.

Property P2 *(order irrelevance):*

$HWM(F_1, ..., F_K) = HWM(G(F_1, ..., F_K))$, where $G(\cdot)$ is any perturbation of the order of its entries.

Property P3 *(consistency):*

$\lim_{n_1,...,n_K \to \infty} HWM\left(F_{1,n_1}, ..., F_{K,n_K}\right) = HWM(F_1, ..., F_K)$.

3.2 Computation

Computing the K-sample HWM index involves again the construction of the underlying d-d plot. In case the p-p plot is projected on the average probabilities, the coordinates of the corresponding d-d plot for any $z_i \in \mathbf{z}$

are:

$$p(z_i) = \frac{1}{K}\sum_{j=1}^{K} F_j(z_i), \tag{8}$$

$$d(z_i) = \sqrt{\sum_{j=1}^{K} \left(F_j(z_i) - p(z_i)\right)^2}, \tag{9}$$

$i = 1, ..., m$, $m \leq \sum_{j=1}^{K} n_j$, and $(p(0), d(0)) \equiv (0, 0)$. Obviously Proposition 1 applies such that the K-sample HWM index can be computed as follows:

$$HWM(F_{1,n_1}, ..., F_{K,n_K}) = PS(K)S(n_1..., n_K) \times \sum_{i=1}^{m+l} \left\{ \frac{(p_i^* - p_{i-1}^*)}{2} \left(2\max\{d_i^*, d_{i-1}^*\} - \left|d_i^* - d_{i-1}^*\right|\right) \right\}, \tag{10}$$

where the notation of Proposition 1 applies.

3.3 Hypotheses testing

An exact formulation is not known for the finite sample distribution of any of the existing K-sample EDF tests with $K > 2$ (Kiefer, 1959). We also have not been able to identify an exact formulation along the lines of Proposition 2 for the distribution of the K-sample HWM index. However, as it is distribution free under assumptions A1* – A2*, we can simulate its distribution. The concomitant critical percentiles are in Tables 10 through 13 in Appendix B for $K = 3, ..., 15$ for sample sizes up to the point of convergence.[7]

4 GDP growth rates across G7 countries

The performance of the HWM test ultimately must be measured against its ability to discriminate between samples that are not drawn from the same

[7] Do note that ties can have a non-negligible effect on the distribution of the K-sample HWM index for $K > 2$, even if the underlying population distributions consist of many classes. This is because the probability of a between-sample tie increases with K relative to the probability of a within-sample tie.

Country	period	mean	st. dev.	min.	max.
Italy (IT)	1970 – 2006	2.3	1.9	-2.0	6.5
United Kingdom (UK)	1948 – 2006	2.5	1.8	-2.1	7.1
Germany (DE)	1960 – 2006	2.8	2.6	-1.3	13.2
United States (US)	1948 – 2006	3.4	2.4	-1.9	8.7
France (FR)	1950 – 2006	3.5	2.0	-1.0	8.5
Canada (CA)	1948 – 2006	3.9	2.4	-2.9	9.5
Japan (JP)	1955 – 2006	4.7	4.2	-5.1	13.1

Table 4: Annual GDP growth rates of all G7 countries, taken from IMF (2007).

population distribution. As none of the existing EDF tests dominates all other tests under all circumstances, in principle all tests must be considered at all times. However, in some cases a particular test is known to have the best power. For instance, when samples are drawn from distributions with different tails, the AD test outperforms all alternatives as it puts more weight on extremities. The HWM test on the other hand is expected to outperform all other tests when the p-p plot remains relatively close to the diagonal because it is the only EDF test that assigns equal weight to all distances between the respective distributions.

As an illustration consider the economic growth rates of all G7 countries from the mid twentieth century up until 2006 (see Table 4; these data are from IMF, 2007). The samples are unbalanced and ties are absent. The seven countries can be arranged in three groups: those with an average growth rate (i) between 2 and 3 percent (Italy, the United Kingdom, and Germany), (ii) between 3 and 4 percent (United States, France, and Canada), and (iii) above 4% (Japan). The distributions of growth rates for all countries are depicted in Figure 3. These distributions differ more clearly between some countries than others, and EDF tests can be used to determine whether these differences are statistically significant.

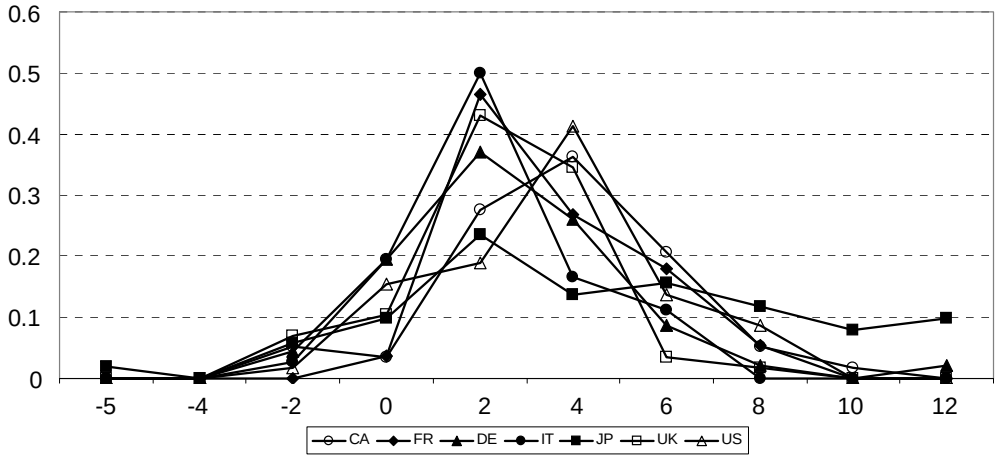

Figure 3: Pdfs of economic growth rates for all G7 countries from the mid 20^{th} century up to 2006.

Note that the right tails of the distributions of Italy and Japan stand out in particular: the one of Italy clearly lies below all other distributions while that of Japan clearly is above all the others. We therefore compare the performance of the HWM index with the alternative EDF test that is specifically designed to deal with this situation, the AD test. Including either Japan or Italy in a group of countries is expected to be picked up most quickly by the latter. On the other hand, the HWM test should outperform the AD test whenever a country is added to a group of countries that have comparable distributions of growth rates.

The results of the two tests are in Table 5.[8] Both tests reject equally strongly the equality of economic growth rates across all G7 countries. This holds also for all possible subsets of 5 or 6 countries (not shown in the table). Further, both tests do not reject at any of the significance levels considered the equality of growth rates within groups (i) and (ii). There are differences

[8] We use the AD test in modified form. In the notation of Scholz and Stephens (1987): $T = \left(A^2 - K + 1\right)/\sigma_N$, where σ_N and A^2 respectively are taken from their formulae (4) and (7).

# samples	comparison	HWM	AD
7	CA, DE, FR, IT, JP, UK, US	1.749****	8.055****
4	CA, DE, FR, US	0.916**	1.736*
	CA, FR, IT, US	1.242****	4.035****
	CA, FR, JP, US	0.854*	2.879***
	CA, FR, UK, US	1.110****	3.857****
	CA, DE, IT, UK	1.250****	5.877****
	DE, FR, IT, UK	0.915*	2.511***
	DE, IT, JP, UK	1.053****	6.145****
	DE, IT, UK, US	0.945**	3.380****
3	CA, DE, US	0.848***	2.215**
	CA, FR, US	0.474	-0.170
	CA, UK, US	1.076****	5.063****
	CA, IT, UK	1.236****	7.666****
	DE, IT, UK	0.474	0.018
	FR, IT, UK	0.891***	3.615****
2	CA, FR	0.375	0.329
	CA, US	0.256	-0.500
	DE, FR	0.517*	1.394*
	DE, IT	0.275	-0.444
	DE, UK	0.316	0.228
	FR, US	0.279	-0.254
	IT, UK	0.385	0.164

Table 5: Annual growth rate comparisons for G7 countries. Note: *, **, ***, and **** refers to H0 being rejected at the 10, 5, 2.5, and 1 percent significant level respectively.

though. The AD tests rejects H_0 at higher significance levels than the HWM test in 4 cases: {CA, FR, JP, US}, {DE, FR, IT, UK}, {DE, IT, UK, US}, and {FR, IT, UK}. As expected, in all these cases either Japan or Italy is included in the comparison. On the other hand, the HWM test rejects H_0 at higher significance levels than the AD-test in 2 cases: {CA, DE, FR, US}, and {CA, DE, US}. Here the difference between the group members (i.e. CA, FR, and US) and the other country included (DE) is relatively small, such that the HWM test has more power.

5 Discussion and conclusions

The quantification of the p-p plot we propose assigns equal weight to all distances between the respective distributions. This allows us to derive the finite sample distribution of the concomitant EDF test in case of two balanced samples absent ties. The formula for the exact computation of the HWM index, which is applicable under all circumstances, allows for simulations in cases where this distribution is not known. Both ties and unbalanced samples appear to affect the distribution of the HWM index only mildly. We have not been able to derive the analytical expression for the distribution of the K-sample HWM index. But with the aid of the d-d plot we can calculate it again under all circumstances, which prompts us to provide relevant significance tables for $K = 3, ..., 15$. As none of the existing EDF tests, including the HWM test, outperforms all other tests under all circumstances, they should all be considered simultaneously so as to avoid type-II errors as much as possible. It is our contribution to have added one member to this family.

References

[1] Anderson, T. W. and D. A. Darling (1952), "Asymptotic theory of certain goodness of fit criteria based on stochastic processes", *Annals of Mathematical Statistics* **23**: 193 – 212.

[2] Bamber, D. (1975), "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph", *Journal of Mathematical Psychology* **12**: 387 – 415.

[3] Cramér, H. (1928), "On the composition of elementary errors II: statistical applications", *Skandinavisk Aktuarietidskrift* **11**: 141 – 180.

[4] Dufour, J. M. (1995), "Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics", Technical report C.R.D.E., Universite de Montreal.

[5] Dufour, J. M. and J. F. Kiviet (1998), "Exact inference methods for first-order autoregressive distributed lags models", *Econometrica* **66**: 79 – 104.

[6] Fisz, M. (1960), "On a result by M. Rosenblatt concerning the von Mises-Smirnov test", *Annals of Mathematical Statistics* **31**: 427 – 429.

[7] Girling, A. J. (2000), "Rank statistics expressible as integrals under P-P-plots and receiver operating characteristics curves", *Journal of the Royal Statistical Society* **B 62**: 367 – 382.

[8] Holmgren, E. B. (1995), "The p-p plot as a method of comparing treatment effects", *Journal of the American Statistical Society* **90**: 360 – 365.

[9] IMF (2007), *International Financial Statistics*, http://www.imf.org.

[10] Johnson, B. McK. and T. Killeen (1983), "An explicit formula for the C.D.F. of the L_1 norm of the Brownian bridge", *The Annals of Probability* **11**: 807 – 808.

[11] Kiefer, J. (1959), "K-sample analogues of the Kolmogorov-Smirnov and Cramer-v.Mises tests", *Annals of Mathematical Statistics* **30**: 420 – 447.

[12] Kolmogorov, A. N. (1933), "Sulla determinizazione empirica delle leggi di probabilita", *Giornale dell 'Istituto Italiano Attuari* **4**: 1 – 11.

[13] Kuiper, N. H. (1960), "Tests concerning random points on a circle", Koninklijke Nederlandse Akademie van Wetenschappen, The Netherlands.

[14] Rosenblatt, M. (1952), "Limit theorems associated with variants of the von Mises statistic", *Annals of Mathematical Statistics* **23**: 617 – 623.

[15] Scholz, F. W. and M. A. Stephens (1987), "K-sample Anderson-Darling tests", *Journal of the American Statistical Association*, **82**: 918 – 924.

[16] Schmidt, F. and M. Trede (1995), "A distribution free test for the two sample problem for general alternatives", *Computational Statistics & Data Analysis* **20**: 409 – 419.

[17] Smirnov, N. V. (1939), "On the deviation of the empirical distribution function", *Rec. Math.* [*Mathematicheskii Sbornik*] *N.S.* **6**: 3 – 26.

[18] Von Mises, R. (1931), *Wahrscheinlichkeitsrechnung*, Vienna: Deuticke.

[19] Wilk, M. B. and R. Gnanadesikan (1968), "Probability plotting methods for the analysis of data", *Biometrika* **55**: 1 – 17.

6 Appendix A Proofs

6.1 Proofs of properties

Proof of Property P1 Let $[a,b]$ be the common support of $\widetilde{F}_j(z)$, $j = 1,...,K$. $HWM(F_1,...,F_K) = 0 \Longleftrightarrow$ because of the continuity of both $\widetilde{F}_j(z)$ and $\widetilde{p}(z)$: $\sqrt{\sum_{j=1}^{K}\left(\widetilde{F}_j(z)-\widetilde{p}(z)\right)^2} = 0\ \forall\ z \in [a,b] \Longleftrightarrow \widetilde{F}_1(z) = \widetilde{p}$,...,$\widetilde{F}_K(z) = \widetilde{p}$, $\forall\ z \in [a,b] \Longleftrightarrow F_1(z) = ... = F_K(z)$, $\forall\ z \in [a,b]$.

QED

Proof of Property P2 $HWM(F_1,...,F_K) \equiv PS(K)S(n_1,...,n_K)\times$

$$\int_{-\infty}^{\infty}\sqrt{\sum_{i=1}^{K}\left(\widetilde{F}_i(x)-\widetilde{p}(x)\right)^2}d\widetilde{p}(x)$$

$$= PS(K)S(n_1,...,n_K)\int_{-\infty}^{\infty}\sqrt{G\left(\sum_{i=1}^{K}\left(\widetilde{F}_i(x)-\widetilde{p}(x)\right)^2\right)}d\widetilde{p}(x)$$

$=HWM(G(F_1,...,F_K))$, where $G(\cdot)$ is any perturbation of the order of its entries.

QED

Proof of Property P3 By application of Slutsky's theorem:

$$\lim_{n_1,...,n_K\to\infty}\int_{-\infty}^{\infty}\sqrt{\sum_{i=1}^{K}\left(\widetilde{F}_{i,n_i}(x)-\widetilde{p}(x)\right)^2}d\widetilde{p}(x)$$

$$=\int_{-\infty}^{\infty}\sqrt{\sum_{i=1}^{K}\left(lim_{n_1,...,n_K\to\infty}\widetilde{F}_{i,n_i}(x)-lim_{n_1,...,n_K\to\infty}\widetilde{p}(x)\right)^2}d\widetilde{p}(x)$$

$$=\int_{-\infty}^{\infty}\sqrt{\sum_{i=1}^{K}\left(\widetilde{F}_i(x)-\widetilde{p}(x)\right)^2}d\widetilde{p}(x).$$

QED

6.2 Proofs of Lemmata

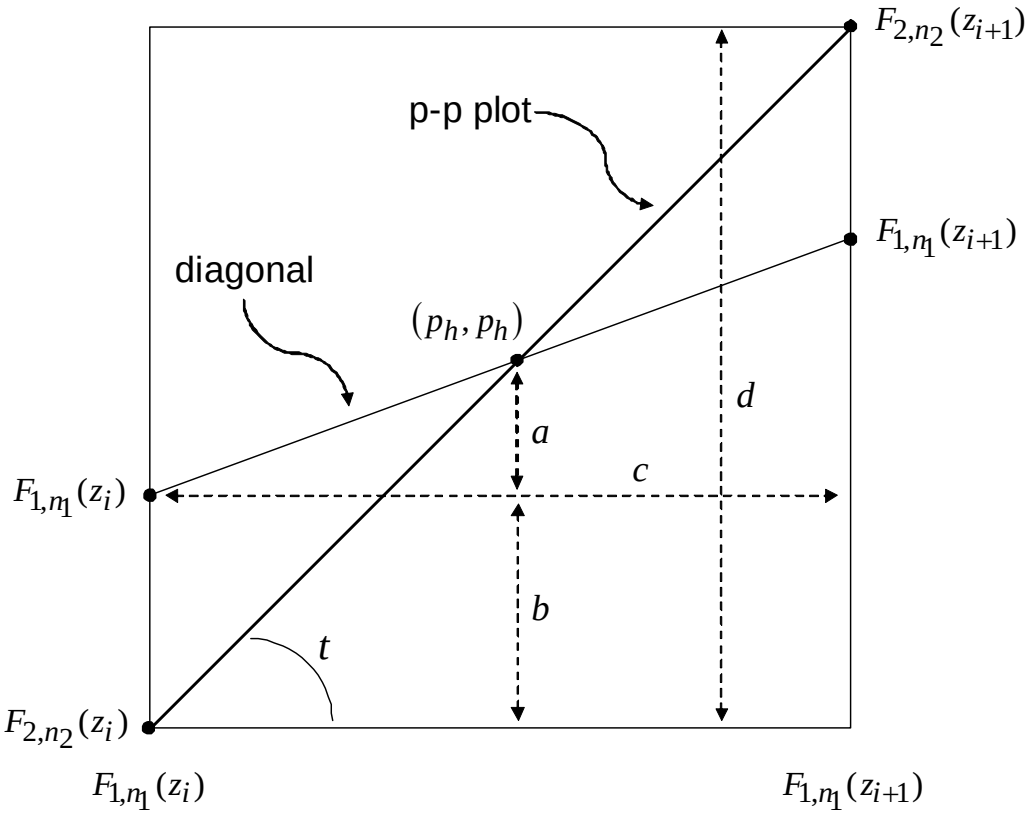

Figure 4: A p-p plot cutting the diagonal "from below" at point (p_h, p_h).

Proof of Lemma 1 A p-p plot that cuts the diagonal from below at probability p_h between observations z_i and z_{i+1} (i.e. $F_{1,n_1}(z_i) \geq F_{2,n_2}(z_i)$ and $F_{1,n_1}(z_{i+1}) \leq F_{2,n_2}(z_{i+1})$) is illustrated in Figure 4. The diagonal runs from $(F_{1,n_1}(z_i), F_{1,n_1}(z_i))$ to $(F_{1,n_1}(z_{i+1}), F_{1,n_1}(z_{i+1}))$ while the p-p plot connects $(F_{1,n_1}(z_i), F_{2,n_2}(z_i))$ with $(F_{1,n_1}(z_{i+1}), F_{2,n_2}(z_{i+1}))$. Let $a = p_h - F_{1,n_1}(z_i)$, $b = F_{1,n_1}(z_i) - F_{2,n_2}(z_i)$, $c = F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i)$, and $d = F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i)$. From $TAN(t) = d/c = (a+b)/a$ we obtain that $a = bc/(d-c)$. Hence, p_h follows. A similar argument applies in case the diagonal is cut from above.

QED

Proof of Lemma 2 Define a "step" on the grid of the sample p-p plot as an increase of the p-p plot of length $1/n$ in either the horizontal or vertical direction. Under A1-A2 the smallest value of the sample HWM index obtains when each next step of the p-p plot is vertical (horizontal) after a horizontal (vertical) step. In that case the p-p plot consists of n triangulars yielding as HWM index: $S(n,n)n/2n^2 = 1/\sqrt{8n}$. The index obtains its largest value when n consecutive steps are either vertical or horizontal, yielding as index value: $S(n,n)/2 = \sqrt{n/8}$. The smallest difference between two sample HWM index values is the surface of a square with length $1/n$ corrected for sample size: $S(n,n)/n^2 = 1/\sqrt{2n^3}$. Hence, the number of distinct sample HWM index values is $1 + \left(\sqrt{n/8} - 1/\sqrt{8n}\right) \Big/ \left(1/\sqrt{2n^3}\right) = 1 + n(n-1)/2$.

QED

Proof of Lemma 3 First note that $HWM_n(j)$, $j = 1, ..., 1 + n(n-1)/2$ is decreasing in j. It then follows that $HWM_n(1) = \sqrt{n/8}$ is the largest value of the sample HWM index, and that $HWM_n(1 + n(n-1)/2) = 1/\sqrt{8n}$ is the smallest sample HWM index value. As $HWM_n(j+1) - HWM_n(j) = 1/\sqrt{2n^3}$, the lemma follows.

QED

6.3 Proofs of propositions

Proof of Proposition 1 The surface between the p-p plot and the diagonal equals the surface below the d-d plot multiplied by $\sqrt{K}$. The surface below the d-d plot between (p^*_{i-1}, d^*_{i-1}) and (p^*_i, d^*_i) equals $(p^*_i - p^*_{i-1})d^*_{i-1} + (p^*_i - p^*_{i-1})(d^*_i - d^*_{i-1})/2$ if $d^*_i \geq d^*_{i-1}$, and $(p^*_i - p^*_{i-1})d^*_i + (p^*_i - p^*_{i-1})(d^*_{i-1} - d^*_i)/2$ if $d^*_i \leq d^*_{i-1}$. The proposition then follows.

QED

Proof of Proposition 2 We first characterize all different p-p plots for sample size n, we then order all these p-p plots according to the corresponding value of the HWM index, and we conclude with combining these HWM index values with their relative frequency of occurrence.

To identify all different p-p plots for sample size n we examine what happens when the sample size increases with one observation. Increasing the sample size from $n = 0$ tot $n = 1$ creates two possible p-p plots: one going from (0,0) to (1,1) via (0,1) and one going from (0,0) to (1,1) via (1,0). This is illustrated in panel a of Figure 5. Adding another observation creates six p-p plots in total. As illustrated in panel b, three of these p-p plots go through point a, and three run through point b. Panel c applies when going from $n = 2$ to $n = 3$. As of points c and f there are four possible continuations of the p-p plot, while there are three possible continuations from points d and e onwards, leading to 20 possible p-p plots in total, and so on.

To capture the recursive pattern in which the set of possible p-p plots evolves when the sample size increases, we identify *border points*: all points on the grid of the p-p plot that are one step short of one, and only one probability being unity. Point e in panel c is a border point, while it is

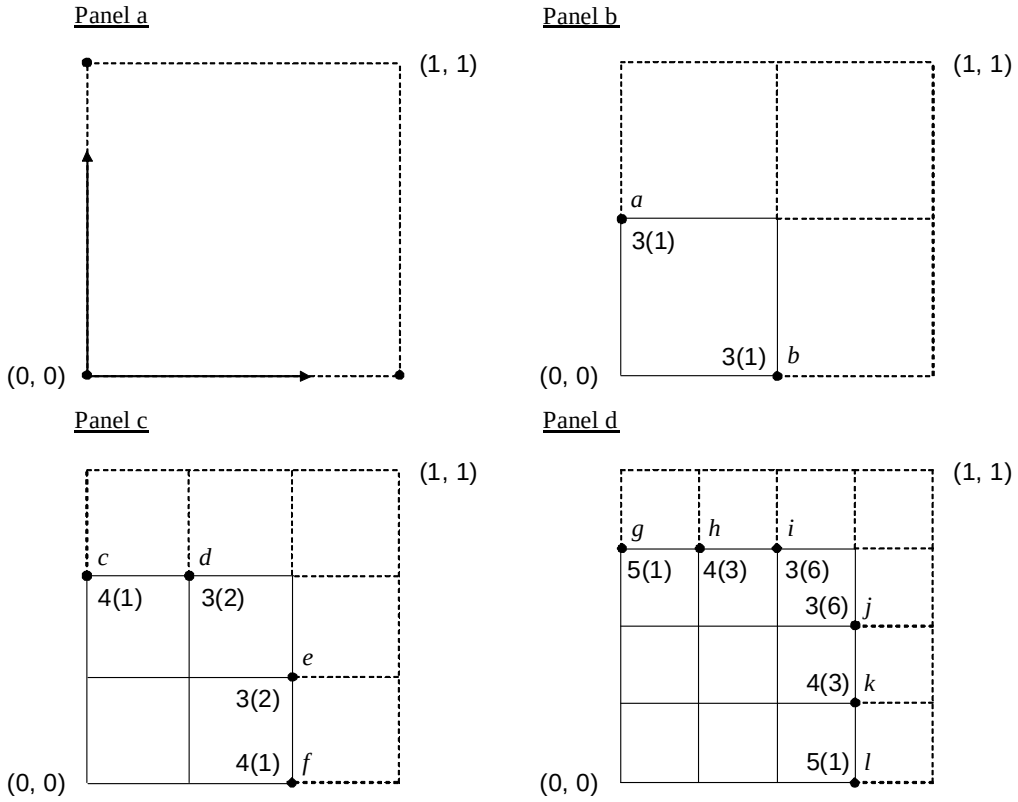

Figure 5: P-p plots, border points, border numbers and history numbers; panels a through d respectively refer to n going from 0 to 1, from 1 to 2, from 2 to 3, and from 3 to 4.

not in panel d. Note that all p-p plots emanating from border points with sample size n necessarily pass through any of the border points for sample size $n+1$. To reveal the pattern in the development of possible p-p plots it suffices therefore to keep track of what happens at the border points.

Let $x(n)$ be a border point given sample size n and let $X(n)$ denote the set of all border points for this sample size. For each border point we introduce two numbers. The *border number*, $bn(x(n))$, is the number of border points that can be reached when the sample size increases with one observation. For example, as of point d in panel c three border points can be reached in panel d: h, i and j, yielding $bn(d) = 3$. At the same time, as of point d there are three possible routes to get at (1,1). Indeed, $bn(x(n))$ coincides with the number of different p-p plot continuations at $x(n)$. Border numbers are uniquely related to the distance between the concomitant p-p plot and the diagonal: the higher is the border number, the larger is the difference between the diagonal and the p-p plot.

The *history number*, $hn(x(n))$, is the number of distinct routes that a p-p plot can have taken to reach $x(n)$ from any $x(n-1)$ onwards. For instance, point h in panel d can be reached in three different ways from two border points in panel c: twice from c and once from d. This yields $hn(h) = 3$. History numbers are one-to-one related to the relative frequency of the concomitant border point: the higher is the history number, the larger is the probability that the particular border point is part of a p-p plot.

In Figure 5 both the border numbers and history numbers are depicted whereby the history numbers are in brackets. For any $n > 1$ the number of distinct p-p plots then equals:

$$\Omega(n) = \sum_{x(n) \in X(n)} bn(x(n)) hn(x(n)).$$

This corresponds to $\sum_i \sum_j a_{ij}(n)$ in the notation of Proposition 2.

To keep track of all possible p-p plots when n expands it suffices to identify the recursive pattern in the development of both border numbers and history numbers when the sample size increases. For that we introduce a logical tree as in Figure 6. It combines all border points with the same border number, split up according to the border points they emanate from. Note that there is a one-to-one correspondence between this tree and the p-p plots in Figure 5. Going from $n = 0$ to $n = 1$ creates two possible p-p plots such that the history number is 2 at the start of the tree. Adding another observation creates two border points, a and b, that both can be reached in one way only. The logical tree then groups the border points with the same border number and history number. In case $n = 3$ there are four border points, two of which have border number 3 (d and e), and two that have border number 4 (c and f). Points d and e can be reached from both points a and b, while points c and f can be reached only from points a and b respectively. Border number 3 can thus emerge from four different p-p plots while there are two distinct p-p plots that lead to border number 4. The tree in Figure 6 splits up accordingly.

Going then from $n = 3$ to $n = 4$ nicely illustrates that points on the grid of the p-p plot can enter Figure 6 at more than one position. Consider for instance point i in panel d of Figure 5. That could have been reached through d or e, yielding $hn(i|\, d, e) = 4$. It could also have been reached through c or f, yielding $hn(i|\, c, f) = 2$. Hence the history number 6 at point i (Figure 5, panel d). Obviously, the probability that a p-p plot that runs through i passes through d or e is twice as large as the probability that it passes through c or f.

Considering then the development of border numbers when the sample

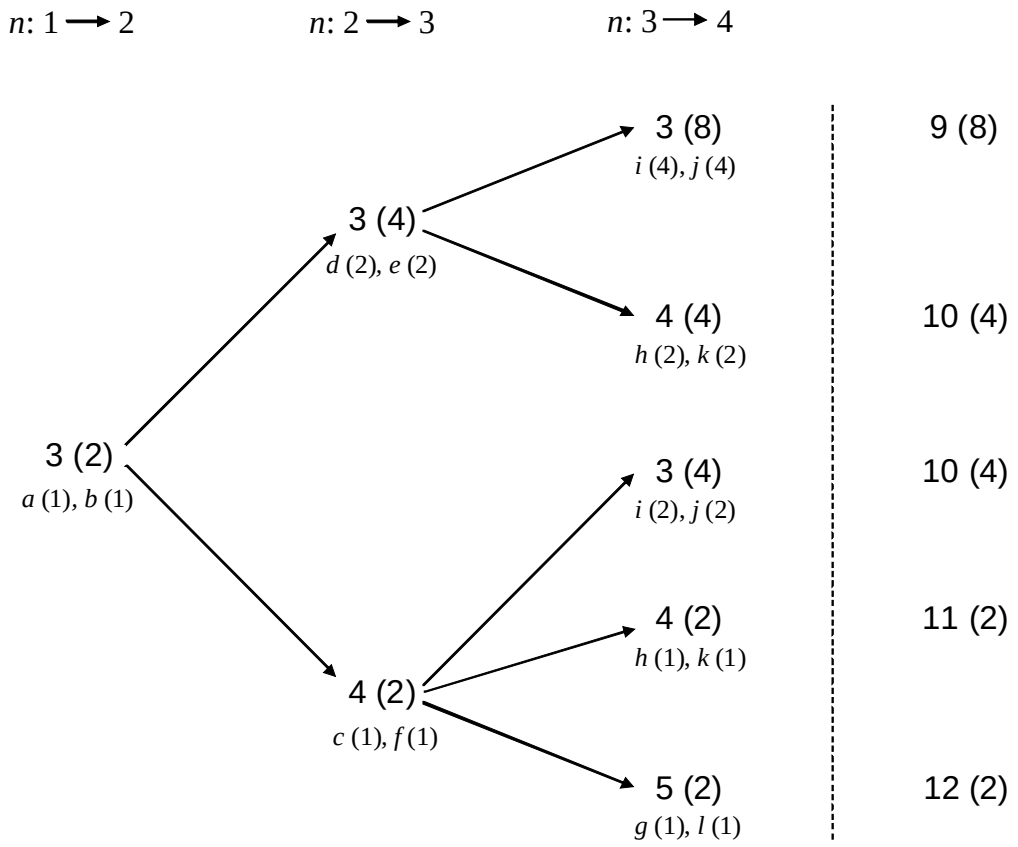

Figure 6: Logical tree for the evolution of possible p-p plots as the sample size increases.

size increases, Figure 6 suggests that border numbers for sample size n always give rise to the same logical sequence of border numbers for sample size $n+1$. In particular, $bn(x(n))$ splits into $bn(x(n))-1$ border numbers with values $\{3,4,...,bn(x(n))+1\}$ when the sample size increases from n to $n+1$. For example, border number 3 is split into $\{3,4\}$ while border number 4 evolves into $\{3,4,5\}$. This is due to a recursive pattern indeed. In general, from any border point $x(n)$ all border points with respective border numbers $\{3,4,...,bn(x(n))\}$ can be reached at least once. In addition, there is one border point with border number $bn(x(n))+1$ that can be reached as well. These possibilities yield the recursive pattern in the evolution of border numbers in Figure 6 as the sample size increases.

Regarding the development of history numbers, first note that in Figure 6 the history numbers of all border points that enter the tree at the same position are summed up. This is because these entries give rise to p-p plots that yield the same value for the HWM index. Then observe that from any border point $x(n)$ all border points with respective border numbers $\{4,5,...,bn(x(n))+1\}$ can be reached once, and only once. Accordingly, the history numbers do not change. For instance, the history number for point h insofar the p-p plot runs through point c is the same as the history number at point c, while it is equal to the history number of point d for all p-p plots that run through point h via point d. That is, $hn(h|\,c) = hn(c) = 1$, and $hn(h|\,d) = hn(d) = 2$. The only exception to this rule are the border points with border number 3. These can be reached from two different border points that have the same border number and history number. And because these border points are grouped together in the logical tree in Figure 6, the history number of border points with border number 3 is twice the history number of the border point they emanate from. For instance, as of point c border

points g, h, i, and j can be reached, with respective border numbers 5, 4, 3, and 3. Border points i and j can also be reached from point f, while points g and h cannot. Therefore, in the bottom branch of Figure 6 we have that $hn(i|\, c, f) = hn(j|\, c, f) = hn(c) + hn(f) = 2$, while $hn(h|\, c, f) = hn(c) = 1$, and $hn(g|\, c, f) = hn(f) = 1$.

The recursive pattern in the evolution of border numbers and history numbers is formalized in the matrix notation of Proposition 2. In particular, the matrix BM contains the development factors of the history numbers, and the history numbers themselves are in HM. For example, for $n = 3$ we have for $BM(3)$:

$$\begin{array}{c} \\ 5 \\ 4 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 6 & 5 & 4 & 3 \end{array} \\ \left[\begin{array}{cccc} 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 \end{array}\right] \end{array},$$

where rows and columns respectively refer to the border numbers for $n = 3$ and $n = 4$, while $HM(3)$ boils down to:

$$\begin{array}{c} \\ 12 \\ 11 \\ 10 \\ 9 \end{array} \begin{array}{c} \begin{array}{ccc} 5 & 4 & 3 \end{array} \\ \left[\begin{array}{ccc} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 4 & 4 \\ 0 & 0 & 8 \end{array}\right] \end{array},$$

where the columns refer to the possible border numbers for $n = 3$, and where the rows refer to the border sums for $n = 3$, as explained next.

Recall that $bn(x(n))$ is one-to-one related to the distance between the diagonal and the p-p plot that runs through $x(n)$. This implies that the sum over all border numbers passed through to complete the p-p plot is uniquely related to the surface between the p-p plot and the diagonal. These *border sums* are included as the last column in Figure 6, with the related history numbers in brackets. Border sums are by definition uniquely related to the different values of the HWM index, in this case for $n = 3$. There are four

distinct border sums that corresponds to the four different values of the HWM index as given in Lemma 3: $(1, 7/9, 5/9, 3/9)\sqrt{3/8}$. Their respective relative frequencies then equal the probabilities that the concomitant border sums emerge, which follow from their history numbers: $(2/20, 2/20, 8/20/8/20)$.

QED

7 Appendix B Critical percentiles

7.1 $K = 2$, unbalanced samples, no ties

7.2 $K = 3, ..., 15$, balanced samples, no ties

n	3	4	5	6	7	8	9	10	11
3	0.612								
4	0.546	0.530							
5	0.502	0.522	0.538						
6	0.511	0.516	0.495	0.529					
7	0.518	0.513	0.512	0.514	0.515				
8	0.492	0.510	0.510	0.501	0.518	0.500			
9	0.528	0.508	0.498	0.498	0.501	0.505	0.511		
10	0.516	0.507	0.511	0.506	0.505	0.502	0.508	0.514	
11	0.501	0.506	0.504	0.508	0.512	0.509	0.506	0.502	0.494
12	0.516	0.505	0.501	0.500	0.508	0.506	0.504	0.506	0.502
13	0.503	0.505	0.512	0.498	0.504	0.502	0.503	0.512	0.504
14	0.495	0.504	0.498	0.505	0.496	0.504	0.508	0.513	0.504
15	0.509	0.503	0.512	0.506	0.502	0.505	0.502	0.504	0.498
16	0.497	0.503	0.508	0.500	0.504	0.505	0.500	0.500	0.505
17	0.507	0.501	0.501	0.510	0.508	0.508	0.508	0.502	0.512
18	0.510	0.503	0.506	0.511	0.506	0.498	0.507	0.507	0.505
19	0.499	0.503	0.508	0.506	0.500	0.502	0.499	0.499	0.506
20	0.501	0.502	0.505	0.507	0.510	0.509	0.501	0.503	0.513
n	12	13	14	15	16	17	18	19	20
12	0.510								
13	0.500	0.505							
14	0.508	0.509	0.499						
15	0.502	0.508	0.510	0.505					
16	0.506	0.506	0.501	0.500	0.508				
17	0.502	0.502	0.501	0.504	0.507	0.499			
18	0.505	0.498	0.501	0.508	0.495	0.502	0.500		
19	0.499	0.506	0.506	0.503	0.506	0.504	0.502	0.499	
20	0.505	0.502	0.502	0.505	0.511	0.500	0.495	0.509	0.506

Table 6: Simulated critical values at percentile 90 for unbalanced samples absent ties. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. Note that the joint number of sample entries must be at least 6 for percentile 90 to be defined properly.

n	3	4	5	6	7	8	9	10	11
4	0.655	0.619							
5	0.593	0.596	0.601						
6	0.550	0.581	0.606	0.577					
7	0.587	0.570	0.581	0.599	0.592				
8	0.577	0.587	0.596	0.582	0.587	0.594			
9	0.583	0.601	0.578	0.586	0.583	0.575	0.589		
10	0.608	0.592	0.584	0.581	0.583	0.580	0.595	0.581	
11	0.582	0.584	0.590	0.584	0.593	0.587	0.578	0.583	0.591
12	0.570	0.577	0.595	0.583	0.584	0.593	0.588	0.585	0.590
13	0.580	0.582	0.599	0.572	0.578	0.578	0.581	0.588	0.586
14	0.569	0.576	0.576	0.580	0.584	0.584	0.594	0.587	0.596
15	0.580	0.565	0.590	0.587	0.582	0.590	0.589	0.588	0.578
16	0.569	0.587	0.586	0.588	0.591	0.586	0.583	0.587	0.580
17	0.579	0.582	0.573	0.585	0.589	0.593	0.587	0.587	0.592
18	0.584	0.578	0.593	0.589	0.588	0.583	0.590	0.585	0.587
19	0.583	0.579	0.577	0.581	0.576	0.590	0.579	0.579	0.582
20	0.569	0.589	0.585	0.579	0.591	0.589	0.584	0.587	0.593
n	12	13	14	15	16	17	18	19	20
12	0.578								
13	0.591	0.581							
14	0.590	0.586	0.594						
15	0.591	0.594	0.589	0.590					
16	0.586	0.592	0.586	0.591	0.586				
17	0.585	0.575	0.582	0.592	0.585	0.580			
18	0.584	0.577	0.581	0.583	0.576	0.580	0.593		
19	0.583	0.587	0.587	0.585	0.591	0.584	0.580	0.585	
20	0.586	0.583	0.583	0.586	0.589	0.583	0.581	0.598	0.585

Table 7: Simulated critical values at percentile 95 for unbalanced samples absent ties. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. Note that the joint number of sample entries must be at least 7 for percentile 95 to be defined properly.

n	3	4	5	6	7	8	9	10	11
5	0.685	0.671	0.664						
6	0.629	0.645	0.661	0.674					
7	0.656	0.627	0.659	0.642	0.668				
8	0.615	0.663	0.658	0.656	0.656	0.656			
9	0.639	0.647	0.657	0.650	0.646	0.657	0.642		
10	0.658	0.651	0.657	0.645	0.667	0.659	0.653	0.648	
11	0.641	0.662	0.657	0.657	0.673	0.660	0.663	0.645	0.649
12	0.645	0.650	0.658	0.653	0.651	0.662	0.658	0.662	0.655
13	0.661	0.673	0.658	0.635	0.648	0.655	0.660	0.660	0.655
14	0.636	0.661	0.658	0.645	0.661	0.665	0.662	0.670	0.662
15	0.650	0.630	0.658	0.646	0.655	0.662	0.659	0.653	0.660
16	0.635	0.643	0.647	0.653	0.650	0.659	0.659	0.660	0.640
17	0.642	0.635	0.643	0.661	0.646	0.656	0.660	0.655	0.670
18	0.653	0.653	0.659	0.648	0.659	0.654	0.650	0.662	0.673
19	0.635	0.646	0.644	0.643	0.655	0.664	0.658	0.649	0.651
20	0.646	0.662	0.660	0.657	0.667	0.663	0.651	0.658	0.666
n	12	13	14	15	16	17	18	19	20
12	0.646								
13	0.657	0.671							
14	0.664	0.656	0.675						
15	0.666	0.670	0.663	0.663					
16	0.668	0.662	0.659	0.661	0.663				
17	0.658	0.654	0.654	0.661	0.664	0.661			
18	0.671	0.657	0.657	0.657	0.647	0.655	0.667		
19	0.662	0.669	0.651	0.655	0.666	0.654	0.658	0.662	
20	0.658	0.659	0.656	0.664	0.662	0.651	0.658	0.672	0.656

Table 8: Simulated critical values at percentile 97.5 for unbalanced samples absent ties. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. Note that the joint number of sample entries must be at least 8 for percentile 97.5 to be defined properly.

n	3	4	5	6	7	8	9	10	11
5		0.745	0.727						
6	0.707	0.710	0.716	0.722					
7	0.725	0.684	0.708	0.728	0.745				
8	0.739	0.714	0.702	0.733	0.729	0.719			
9	0.694	0.740	0.737	0.738	0.740	0.743	0.746		
10	0.709	0.718	0.730	0.742	0.745	0.738	0.747	0.760	
11	0.675	0.740	0.725	0.746	0.766	0.734	0.730	0.728	0.727
12	0.689	0.722	0.720	0.722	0.726	0.730	0.756	0.743	0.744
13	0.701	0.740	0.745	0.727	0.738	0.749	0.753	0.750	0.743
14	0.711	0.724	0.740	0.732	0.727	0.745	0.755	0.759	0.754
15	0.692	0.711	0.736	0.736	0.739	0.742	0.747	0.746	0.740
16	0.702	0.727	0.715	0.718	0.749	0.740	0.717	0.744	0.740
17	0.705	0.714	0.728	0.743	0.720	0.737	0.737	0.738	0.767
18	0.713	0.729	0.725	0.727	0.737	0.735	0.726	0.758	0.760
19	0.696	0.718	0.723	0.731	0.746	0.749	0.730	0.727	0.739
20	0.700	0.730	0.740	0.734	0.755	0.747	0.734	0.762	0.751
n	12	13	14	15	16	17	18	19	20
12	0.731								
13	0.737	0.762							
14	0.741	0.743	0.756						
15	0.757	0.773	0.743	0.736					
16	0.750	0.753	0.734	0.753	0.740				
17	0.741	0.743	0.736	0.747	0.760	0.741			
18	0.758	0.751	0.746	0.750	0.738	0.744	0.750		
19	0.737	0.759	0.736	0.747	0.756	0.747	0.747	0.764	
20	0.745	0.734	0.748	0.742	0.743	0.739	0.737	0.756	0.743

Table 9: Simulated critical values at percentile 99 for unbalanced samples absent ties. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. Note that the joint number of sample entries must be at least 9 for percentile 99 to be defined properly.

n / K	3	4	5	6	7	8	9	10	50	100
3	0.720	0.706	0.706	0.704	0.710	0.696	0.703	0.705	0.692	0.701
4	0.841	0.840	0.843	0.843	0.840	0.841	0.838	0.837	0.839	0.841
5	0.948	0.951	0.951	0.949	0.953	0.952	0.950	0.953	0.950	0.956
6	1.040	1.045	1.049	1.049	1.048	1.045	1.047	1.052	1.045	1.048
7	1.127	1.133	1.133	1.138	1.135	1.131	1.134	1.138	1.136	1.141
8	1.202	1.214	1.214	1.213	1.215	1.208	1.211	1.217	1.212	1.220
9	1.276	1.286	1.289	1.285	1.288	1.285	1.282	1.288	1.287	1.292
10	1.343	1.354	1.357	1.351	1.355	1.352	1.353	1.357	1.357	1.363
11	1.406	1.418	1.421	1.416	1.422	1.418	1.420	1.421	1.428	1.423
12	1.469	1.478	1.483	1.479	1.482	1.481	1.483	1.484	1.491	1.485
13	1.528	1.538	1.539	1.538	1.540	1.540	1.544	1.545	1.549	1.541
14	1.583	1.592	1.595	1.594	1.596	1.598	1.598	1.598	1.604	1.601
15	1.635	1.643	1.648	1.650	1.648	1.650	1.652	1.653	1.660	1.656

Table 10: Simulated values of the HWM index at percentile 90 under A1* - A2*. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.

n / K	3	4	5	6	7	8	9	10	50	100
3	0.753	0.762	0.779	0.773	0.778	0.770	0.774	0.780	0.776	0.773
4	0.886	0.900	0.904	0.909	0.906	0.906	0.904	0.911	0.910	0.915
5	0.993	1.004	1.012	1.016	1.011	1.015	1.013	1.020	1.025	1.020
6	1.087	1.102	1.109	1.109	1.111	1.111	1.109	1.117	1.121	1.122
7	1.177	1.189	1.191	1.198	1.198	1.197	1.199	1.202	1.204	1.211
8	1.254	1.267	1.273	1.273	1.276	1.280	1.278	1.276	1.283	1.294
9	1.327	1.341	1.346	1.346	1.350	1.347	1.350	1.352	1.361	1.365
10	1.395	1.406	1.411	1.414	1.422	1.419	1.417	1.425	1.431	1.430
11	1.454	1.474	1.477	1.480	1.485	1.484	1.484	1.489	1.497	1.494
12	1.518	1.535	1.536	1.539	1.542	1.545	1.545	1.548	1.559	1.551
13	1.576	1.590	1.597	1.598	1.601	1.603	1.603	1.609	1.618	1.611
14	1.631	1.644	1.654	1.655	1.655	1.661	1.660	1.665	1.671	1.667
15	1.684	1.699	1.706	1.712	1.709	1.713	1.713	1.719	1.726	1.720

Table 11: Simulated values of the HWM index at percentile 95 under A1* - A2*. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.

n / K	3	4	5	6	7	8	9	10	50	100
3	0.805	0.820	0.834	0.831	0.841	0.832	0.831	0.840	0.848	0.844
4	0.926	0.947	0.961	0.964	0.965	0.957	0.961	0.967	0.978	0.978
5	1.035	1.052	1.066	1.074	1.068	1.072	1.072	1.083	1.093	1.087
6	1.127	1.149	1.159	1.168	1.166	1.172	1.170	1.175	1.184	1.189
7	1.214	1.236	1.241	1.252	1.260	1.254	1.255	1.257	1.269	1.274
8	1.294	1.311	1.328	1.327	1.330	1.338	1.334	1.336	1.346	1.356
9	1.366	1.385	1.392	1.398	1.405	1.405	1.408	1.410	1.419	1.425
10	1.434	1.455	1.457	1.473	1.476	1.472	1.473	1.480	1.489	1.487
11	1.500	1.521	1.526	1.533	1.540	1.543	1.534	1.544	1.561	1.549
12	1.557	1.580	1.586	1.594	1.603	1.605	1.597	1.604	1.621	1.609
13	1.617	1.637	1.647	1.651	1.658	1.665	1.658	1.663	1.681	1.670
14	1.671	1.692	1.706	1.707	1.707	1.718	1.715	1.716	1.740	1.730
15	1.727	1.747	1.755	1.764	1.761	1.772	1.771	1.773	1.790	1.785

Table 12: Simulated values of the HWM index at percentile 97.5 under A1* - A2*. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.

n / K	3	4	5	6	7	8	9	10	50	100
3	0.839	0.874	0.888	0.896	0.909	0.900	0.906	0.917	0.919	0.930
4	0.968	1.008	1.013	1.024	1.033	1.027	1.037	1.042	1.057	1.055
5	1.082	1.111	1.125	1.135	1.142	1.144	1.138	1.150	1.172	1.169
6	1.174	1.208	1.212	1.225	1.233	1.247	1.244	1.246	1.256	1.254
7	1.259	1.287	1.301	1.310	1.318	1.327	1.329	1.327	1.341	1.341
8	1.333	1.371	1.382	1.394	1.392	1.393	1.406	1.394	1.416	1.422
9	1.412	1.434	1.454	1.464	1.470	1.479	1.470	1.472	1.494	1.490
10	1.477	1.505	1.522	1.532	1.534	1.543	1.532	1.551	1.557	1.547
11	1.545	1.574	1.585	1.600	1.599	1.603	1.600	1.612	1.636	1.623
12	1.609	1.628	1.645	1.658	1.660	1.664	1.667	1.670	1.700	1.686
13	1.665	1.685	1.704	1.714	1.719	1.724	1.728	1.727	1.757	1.740
14	1.721	1.748	1.758	1.771	1.767	1.784	1.784	1.782	1.805	1.805
15	1.783	1.801	1.814	1.821	1.822	1.836	1.836	1.833	1.859	1.860

Table 13: Simulated values of the HWM index at percentile 99 under A1* - A2*. The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.